



PUBVET, Publicações em Medicina Veterinária e Zootecnia.

Aplicação do modelo de limiar utilizando a Inferência Bayesiana no melhoramento animal

Natanael Pereira da Silva Santos¹, José Lindenberg Rocha Sarmento², José Elivalto Guimarães Campelo³, Aurino de Araujo Rego Neto¹, Luiz Antonio Silva Figueiredo Filho⁴, Márcio da Silva Costa¹, Daniel Biagiotti¹, Fabrício Fernandes Guimarães⁵

¹Programa de Pós-Graduação em Ciência Animal da UFPI. Teresina-PI. E-mail: natanaelpss@hotmail.com

²Professor, Doutor do Campus Cinobelina Elvas, Bom Jesus - UFPI

³Professor, Doutor do Campus Petrônio Portela, Teresina - UFPI.

⁴Médico Veterinário, Polo São Raimundo das Mangabeiras - IFMA

⁵Discentes do curso de Medicina Veterinária da UFPI, Bom Jesus – PI.

Resumo

A aplicação de metodologias modernas utilizando o modelo animal é imprescindível para a condução eficiente de qualquer programa de melhoramento. Os procedimentos aplicados para características nas quais a variação contínua é assumida não são adequados para aquelas com variação não contínua. No modelo de limiar assume-se o processo de resposta está associado a uma variável contínua subjacente que apresenta distribuição normal contínua. A representação probabilística de todo o conhecimento incerto é a essência da inferência Bayesiana, seja tal conhecimento relacionado

ao futuro, quantidades observáveis ou parâmetros desconhecidos. Os conceitos de *priori* e *posteriori* são sempre relativos à observação considerada no momento. Assim, objetivou-se com esta revisão apresentar e discutir algumas tendências e aplicações de modelo de limiar sob o enfoque da inferência Bayesiana, aplicado ao melhoramento genético animal.

Palavras-chave: amostragem de Gibbs, componentes de variância, dados categóricos, distribuição discreta, predisposição

Application of threshold model using Bayesian Inference in animal breeding

Abstract

The application of modern methods using the animal model is essential for the efficient conduct of any breeding program. The procedures to characteristics in which continuous variation is assumed are not suitable for those with non-continuous variation. In threshold model assumes that the answering process is associated with an underlying continuous variable which has continuous normal distribution. The probabilistic representation of all knowledge uncertain is the essence of Bayesian inference, such knowledge is related to future observable quantities or unknown parameters. The *priori* and *posteriori* concepts are always relative to the observation at the time considered. Thus, we aimed at with this revision presenting and discussing some trends and applications of threshold model with a focus on Bayesian inference, applied to animal breeding.

Keywords: Gibbs sampling, components of variance, categorical data, discrete distribution, predisposition

Introdução

O conhecimento sobre parâmetros genéticos e fenotípicos é imprescindível para a condução eficiente de qualquer programa de melhoramento animal. Na

última década estes estudos tornaram-se ainda mais reduzidos em virtude da aplicação de metodologias modernas utilizando o modelo animal, sendo um dos maiores entraves à sua utilização, a ausência de informações seguras sobre a matriz de parentesco dos animais.

Da mesma forma, a avaliação da mudança genética em um rebanho é importante para mensurar o grau de eficiência no processo de seleção dos animais. Com esta avaliação é possível verificar se está havendo melhoramento genético no rebanho e se a taxa de ganho é satisfatória. Caso a resposta não seja favorável, é possível identificar as falhas e estabelecer métodos de seleção mais eficientes.

No melhoramento genético animal a pressuposição de distribuição normal dos efeitos aleatórios genéticos e não-genéticos é adotada nos métodos de estimação de componentes de variâncias, fato este que em grandes amostras é fácil de verificar (teorema do limite central). Entretanto, em amostras pequenas, como são as análises a nível de rebanho, a distribuição pode não se aproximar da normal, ou pode se tratar de uma normal contaminada, tornando a análise muito sensível à presença de observações discrepantes, sendo necessária a utilização de metodologias que levem em consideração todas as informações dos animais e o tipo de distribuição dos dados.

Neste contexto, os métodos de Monte Carlo via Cadeias de Markov (MCMC), dentre os quais se destaca a Amostragem de Gibbs, podem ser utilizados como ferramenta na análise de dados, propiciando uma inferência Bayesiana, a partir da qual são obtidas as estimativas dos componentes de variância e parâmetros genéticos, permitindo ainda a incorporação de informações passadas (*a priori*), caso existam, enriquecendo o processo de inferência (FARIA et al., 2007).

Em função disso, objetivou-se com esta revisão apresentar e discutir algumas tendências e aplicações de modelo de limiar sob o enfoque da inferência Bayesiana, aplicado ao melhoramento genético animal.

Característica de Limiar

Contrastando com as características de produção, várias características biológicas de importância econômica na produção animal, como algumas reprodutivas, apresentam distribuição fenotípica descontínua. Entretanto, essas têm distribuição contínua subjacente e, aparentemente, estão fora do contexto da genética quantitativa, mas quando submetidas a análise genética, apresentam-se como se estivessem sob influência de muitos genes, como as características métricas que exibem variações contínuas. Isso significa que, mesmo sob herança poligênica, os fenótipos são classificados dentro de uma ou várias categorias (GIANOLA, 1982).

O conceito de limiar foi originalmente introduzido por Wright (1934). Posteriormente Falconer (1965) descreveu a herança das características de limiar. O princípio do modelo de limiar para características categóricas é baseado na pressuposição que a expressão fenotípica da característica está associada a uma variável contínua subjacente que não é observável. Ou seja, esse modelo assume que a expressão categórica de uma característica, por exemplo, a prolificidade, é uma reflexão de uma 'predisposição' do animal parir.

A escala subjacente é contínua e a escala visível ou observável é descontínua, mas as duas são conectadas pelo limiar ou ponto de descontinuidade. Isso se torna mais claro para características cuja expressão visível pode tomar, apenas, duas formas (SORENSEN; GIANOLA, 2002), por exemplo: parto simples ou múltiplo.

Assim, animais cujos valores fenotípicos na escala subjacente ultrapassam o limiar, aparecerão numa classe, enquanto os animais abaixo desse limiar aparecerão na outra classe. Dessa forma, assume-se que a expressão do fenótipo na escala subjacente é influenciada por muitos fatores genéticos e ambientais que modificam a manifestação do caráter. Vale ressaltar, que dentro de cada segmento da população, os indivíduos diferem marcadamente

no potencial fenotípico e, possivelmente, no potencial genético, mas não no fenótipo. Isso cria problemas na seleção das características de limiar.

Esse problema fica fácil de ser compreendido através do exemplo reportado por Turner e Young (1969), em que os componentes genéticos e de ambiente que atuam sobre o animal podem contribuir para uma potencialidade de nascer 0,3 crias, mas somente zero ou uma cria pode ser expresso no fenótipo. Então, nenhuma cria irá nascer até que o limiar para uma cria seja atingido. Existe um limiar similar para nascer duas crias, outro para três, e assim por diante.

Problemas na análise de dados categóricos

Os procedimentos para análise de dados genéticos descontínuos estão divididos em dois grupos e a distinção entre eles depende de assumir, ou não, a distribuição contínua subjacente. O primeiro grupo inclui métodos que ignoram a natureza discreta dos dados categóricos e a análise é feita utilizando-se o procedimento linear como se os dados fossem contínuos. O segundo envolve o conceito de limiar, ou *threshold*, que assume uma distribuição normal, não observável, subjacente para a variável discreta mensurada.

A Melhor Predição Linear Não Viesada (BLUP), conforme salientado por Thompson (1979) e Hoeschele e Gianola (1988), não seria apropriada para as variáveis categóricas, pois elas violariam várias pressuposições existentes nas metodologias de modelos mistos. Os autores identificaram e discutiram alguns problemas nas análises de dados binomiais com modelo linear. Alguns dos problemas mais críticos são:

- a) as variâncias das características categóricas são heterogêneas, por que dependem do valor genotípico dos 'candidatos' à seleção;

- b) as soluções BLUP não levam em consideração a restrição que a soma das probabilidades de resposta tenha que ser igual a 1 entre todas as categorias de resposta;
- c) a variância genética aditiva na escala observada depende da incidência da característica na população;
- d) há possibilidade que o efeito genético não aditivo esteja presente na escala observada no caso em que toda a variação genética é aditiva na escala subjacente e;
- e) quando a esperança condicional das predições, dado que os valores das mensurações não sejam linear, as propriedades de classificação do melhor preditor linear não parecem ser otimizadas.

Portanto, teoricamente fica evidente que, os procedimentos aplicados para características nas quais a variação contínua é assumida não são adequados para aquelas com variação não contínua, com isso o desenvolvimento de metodologias específicas para análises de dados categóricos veio ao encontro das necessidades dos pesquisadores, pois auxiliou no estudo de características, como a prolificidade, viabilizando sua utilização na avaliação genética do rebanho.

Comparações entre os dois métodos utilizando a estimação de parâmetros genéticos para o número de crias nascidas foram feitas por Jensen (1986). O autor concluiu que as estimativas de herdabilidade e a capacidade de transmissão dessa característica poderiam ser obtidas por qualquer um desses métodos, pois estão associados. Destacou ainda que, a similaridade das estimativas sugere somente uma pequena vantagem para o modelo não linear em termos de aumento na identificação da variância aditiva. Estudos similares, que compararam tais resultados foram realizados por Olesen et al. (1994) e Sousa et al. (2000).

Modelo de Limiar

No modelo de limiar assume-se o processo de resposta está associado a uma variável contínua subjacente que apresenta distribuição normal contínua chamada *Liability* (Predisposição), representada pela letra U , e também a um conjunto de limiares fixos, $t(t_{\min}, t_1, \dots, t_{j-1}, t_{\max})$, $t_1 < t_2 \dots < t_{j-1}$, que divide a linha real dentro de j intervalos correspondentes às categorias de respostas, com $t_0 = -\infty$ e $t_j = +\infty$. A distribuição de U no contexto do modelo multifatorial da genética quantitativa, pode ser assumida como normal, portanto esta variável é considerada como o resultado da combinação linear de pequenos efeitos a partir de alelos em um grande número de *locus*, mais componentes aleatórios e de ambiente.

Como proposto por Gianola e Foulley (1983), para análises de dados categóricos, os dados são organizados em uma tabela de contingência $s \times j$ (Tabela 1), onde as s linhas representam indivíduos ou combinações de níveis de variáveis explanatórias e as j colunas indicam categorias de respostas orientadas mutuamente, exclusivas e exaustivas.

Nessa tabela, n_{at} são os números de unidades experimentais na k ésima categoria resposta $t(t=1, 2, \dots, j)$, sob as condições da a ésima linha ($a=1, 2, \dots, s$). Totais de linhas, n_a ($a=1, \dots, s$) são considerados como fixos. As variáveis aleatórias de interesse são $n_{a1}, n_{a2}, \dots, n_{at}$ para $a=1, \dots, s$.

Os dados na tabela de contingência podem ser representados simbolicamente por uma matriz $j \times s$. $Y = |y_1, y_2, \dots, y_s|$, em que y_a é um vetor $j \times 1$. Tem-se que:

$$y_a = \sum_{r=1}^{n_a} y_{ar}$$

e y_{ar} é um vetor $j \times 1$, tendo 1 na linha correspondente à categoria de resposta da $ar^{ésima}$ unidade experimental e zero nos outros lugares. y tem distribuição conjunta com o vetor de parâmetros θ , sendo a densidade conjunta $p(Y, \theta)$.

Pelo teorema de Bayes tem-se que: $p(\theta|Y) = p(Y|\theta)p(\theta)/p(Y)$

onde, $p(Y)$ é a função de distribuição dos dados (Y). $p(\theta)$ é a densidade a *priori* de θ (representa o conhecimento prévio a respeito dos elementos de θ antes da observação dos dados em Y); $p(Y|\theta)$ é a função de verossimilhança (representa a contribuição de Y para o conhecimento sobre (θ)). Como $p(Y)$ não varia com θ , a densidade a *posteriori* pode ser usada como:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Associado a cada linha da tabela de contingência ($s \times j$), existe um parâmetro de locação (n_a), tal que a variável subjacente para a $q^{ésima}$ unidade experimental na $a^{ésima}$ linha possa ser escrita como:

$$\text{Predisposição} = U_{aq} = n_a + \varepsilon_{aq},$$

em que: $a = 1, \dots, s$ e $q = 1, \dots, n_a$ e $\varepsilon_{aq} \sim N(0, I\sigma^2)$.

Nota-se que a variância residual é tomada como a unidade de medida da escala subjacente. Além disso, ao parâmetro de locação n_a é dada uma estrutura linear, como exemplo a do modelo animal. O desenvolvimento detalhado do método pode ser encontrado em Gianola e Foulley (1983), Sousa (1997) e Mrode (2005).

Tabela 1 – Dados categóricos ordenados e arranjados em tabela de contingência ($s \times j$)

Linha ¹	Categoria da resposta ²						Total fixo ³
	1	2	...	t	...	j	
1	n_{11}	n_{12}	...	n_{1t}	...	n_{1j}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2t}	...	n_{2j}	$n_{2.}$
⋮	⋮	⋮	...	⋮	...	⋮	⋮
a	n_{a1}	n_{a2}	...	n_{at}	...	n_{aj}	$n_{a.}$
⋮	⋮	⋮	...	⋮	...	⋮	⋮
s	n_{s1}	n_{s2}	...	n_{st}	...	n_{sj}	$n_{s.}$

Fonte: Gianola e Foulley (1983)

¹ Uma linha representa um indivíduo ou uma combinação de variáveis

² n_{at} é o número de observações na $t^{\text{ésima}}$ categoria resposta na $a^{\text{ésima}}$ linha

³ $n_{a.} = \sum_{t=1}^j n_{at}$

Teorema de Bayes

A inferência estatística trabalha na presença de observações (y), cujos valores inicialmente incertos são descritos por meio de uma distribuição de probabilidade $p(y/\theta)$. O parâmetro θ é o que interessa ao pesquisador, e tem um sentido exato dentro da inferência Bayesiana, ou seja, as variâncias dos componentes estudados. O pesquisador deve ter alguma informação prévia sobre θ , e esta informação deve ser incorporada à análise.

Ao incorporar uma opinião sobre θ à análise, através de uma densidade de probabilidade $p(\theta)$, está se denominando a densidade *a priori* que recebe este nome por ser a distribuição de probabilidade de θ antes que se observem os dados. O parâmetro θ pode ser um escalar ou um vetor de parâmetros.

A densidade conjunta de um grupo de observações $Y(y_1, \dots, y_n)$ examinada como uma função do parâmetro θ , é denominada função de verossimilhança e é representada por $p(y_1, \dots, y_n)$, sendo n o número de observações.

Se $p(\theta)$ é a densidade a *priori* para θ , então a densidade a *posteriori* de θ é dada pelo Teorema de Bayes, ou seja:

$$p(\theta|Y) = \frac{p(\theta|Y)p(\theta)}{p(Y)}$$

Como $p(Y)$ não varia com θ , a densidade a *posteriori* pode ser dada por:

$$p(\theta|Y) \propto p(y|\theta)p(\theta)$$

em que α representa proporcionalidade.

Pode-se pensar no Teorema de Bayes como um mecanismo de atualização da opinião do estatístico sobre θ e dessa forma, pode ser visto como elemento essencial para a análise Bayesiana, pois toda inferência é feita a partir da densidade a *posteriori*.

A representação probabilística de todo o conhecimento incerto é a essência da inferência Bayesiana, seja tal conhecimento relacionado ao futuro, quantidades observáveis ou parâmetros desconhecidos. Os conceitos de *priori* e *posteriori* são sempre relativos à observação considerada no momento. Dessa forma, a diferença formal entre a inferência Bayesiana e a frequentista é que, para a inferência Bayesiana, o parâmetro θ é variável aleatória, apresentando, então, uma distribuição de probabilidade. Para a inferência frequentista, os parâmetros são valores fixos, não sendo possível atribuir a eles uma distribuição de probabilidade.

Inferência Bayesiana

Métodos Bayesianos podem ser especialmente úteis em problemas complexos ou nas situações em que naturalmente não há conformidade com o cenário clássico, onde muitos problemas genéticos se enquadram dentro deste.

Gianola (1982) introduziu os métodos Bayesianos no melhoramento animal no contexto de característica de limiar e posteriormente Gianola e Fernando (1986), realçaram as possibilidades de explorar essa técnica.

As técnicas Bayesianas foram abandonadas durante as décadas de 80 e 90, por requerer resolução de múltiplas integrais, geralmente complicadas, muitas vezes com uso de métodos numéricos, que podem não ser flexíveis. A descoberta da técnica Monte Carlo via Cadeias de Markov (MCMC), possibilitou a solução para muitos problemas que não foram resolvidos no passado devido à impossibilidade de resolução dessas integrais. Entretanto, novos problemas, relacionados à convergência das cadeias de Gibbs, foram questionados. Felizmente, esses novos problemas são facilmente manuseáveis, particularmente quando a distribuição dos dados é normal (BLASCO, 2001).

Os métodos MCMC são um conjunto de processos iterativos para a geração aproximada de amostras de distribuições multivariadas. Nestes, o conhecimento da variável aleatória gerada só é relevante para a geração da variável seguinte, sendo, portanto, irrelevante na predição das variáveis futuras. Esses métodos se tornaram importante ferramenta computacional na estatística Bayesiana, uma vez que geram cadeias de Markov por meio de simulação iterativa de Monte Carlo, cuja distribuição estacionária é a própria densidade a *posteriori* de interesse (SORENSEN; GIANOLA, 2002). Um método MCMC bastante utilizado é o Algoritmo de Metroplis-Hastings (HASTINGS, 1970), sendo o Amostrador de Gibbs (GEMAN; GEMAN, 1984) um caso especial deste algoritmo que vem sendo empregado na avaliação genética animal.

O Amostrador de Gibbs é uma técnica indireta para gerar variáveis aleatórias a partir de uma distribuição marginal, sem, no entanto, a necessidade de calcular a sua densidade (ou seja, a função densidade de probabilidade marginal desta variável). Essa técnica é relativamente fácil de ser implementada. Dada a função de máxima verossimilhança e as densidades a *priori*, calcula-se a densidade conjunta a *posteriori* dos parâmetros desconhecidos. A partir dessa densidade, obtém-se a distribuição condicional

completa de cada variável, fixando-se as demais variáveis da densidade conjunta. Esse conjunto de densidades condicionais completas permite a implementação do Amostrador de Gibbs (MRODE, 2005).

Esse processo de amostragem utiliza algoritmos destinados à solução de problemas práticos relacionados à integração numérica da função densidade de probabilidade, impossíveis de serem solucionados por métodos analíticos tradicionais, tornando possível obter amostras das distribuições marginais das variáveis aleatórias.

A utilização do método de amostragem de Gibbs, segundo Van Tassel et al. (1995), pode apresentar vantagens sobre outros métodos, dada a sua relativa simplicidade, permitindo obter estimativas precisas e intervalos de confiança para a distribuição *a posteriori* dos componentes de variância, sem o uso de aproximações ou de suposições de normalidade comumente utilizadas, além de poder ser empregado em análises de grandes conjuntos de dados, além da programação do algoritmo de Gibbs não ser complexa. Dessa forma, fica evidente a possibilidade da análise de dados que apresentam distribuição discreta ou distribuição normal dos dados, possibilitando estimação de componentes de variância e covariância em características de interesse econômico, sejam de caráter reprodutivo ou de desempenho produtivo.

Entretanto, na estatística Bayesiana não existe distinção entre variáveis fixas e aleatórias, todas são consideradas como aleatórias. Efeitos considerados fixos no melhoramento animal (rebanho, ano, etc) são tidos, na estatística Bayesiana, como variáveis aleatórias sobre as quais se tem pouco ou nenhum conhecimento *a priori*, ou seja, *a priori* para o pesquisador é indiferente do provável valor destas variáveis (GIANOLA; FERNANDO, 1986).

Análise da Convergência

Teoricamente, o processo iterativo teria fim apenas quando o número de iterações tendesse ao infinito. Na prática, admite-se convergência quando a série alcança um estado de estacionariedade, significando que as condicionais

completas estão suficientemente próximas das distribuições marginais a *posteriori*, o que pode ser verificado com a utilização de diferentes critérios (NOGUEIRA et al., 2003).

Existem diversos métodos que podem ser usados como indicativos de provável convergência (COWLES; CARLIN, 1996; BROOKS; ROBERTS, 1998) e o uso concomitante de vários desses procedimentos é recomendável. Na prática, para as análises típicas em genética e melhoramento animal tem-se adotado os seguintes procedimentos:

- a) rodar uma cadeia "longa" dependendo da complexidade do modelo, algo entre 100.000 e 200.000 ciclos após um período de *burn-in* de 10% da cadeia e salvando a cada 10 ou 20 amostras (GEYER, 1992);
- b) fazer gráficos de traço dos valores das amostras para todos os componentes de variância versus o ciclo (Figura 1). Ou seja, constrói-se o gráfico de alguma função do estado da cadeia (eixo das coordenadas) *versus* o número de iterações (eixo das abscissas). Espera-se que o gráfico do traço se adapte dentro de uma amplitude de valores plausíveis, quando atingirem estacionariedade. O gráfico do traço deve ser exibido para todos os componentes do vetor de parâmetros θ , embora isso não seja feito quando a dimensão do vetor de parâmetros θ é grande. A recomendação de convergência requer que o gráfico do traço de todos os componentes do vetor atinja convergência (NOGUEIRA et al, 2003);

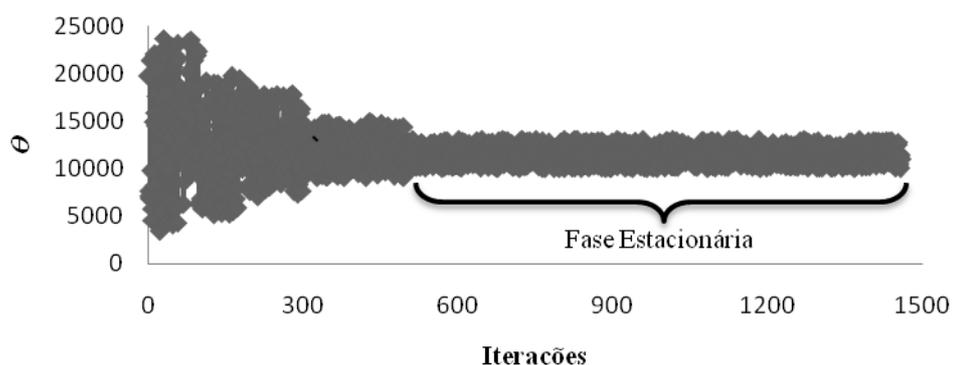


Figura 1 – Representação da análise de convergência da variância pela observação no gráfico do Traço

- c) rodar pelo menos duas outras cadeias de mesmo comprimento da primeira, uma iniciando com valores extremos para cima da média observada dos parâmetros e outra para baixo. Fazer o gráfico de traço concomitante das três cadeias que iniciaram com valores distintos e verificar a convergência para o mesmo espaço amostral, após o período de descarte (CARDOSO, 2008);
- d) realizar o emparelhamento de cadeias de Johnson (1996), que consiste em iniciar duas cadeias com a mesma semente para o gerador de números aleatórios, mas com diferentes valores iniciais. Neste caso, a convergência é indicada pelo perfeito emparelhamento das cadeias após o período de *burn-in* (válida somente para cadeias exclusivas pelo amostrador de Gibbs, sem passos de Metropolis).

Por ser um processo iterativo de amostragem, as amostras sequenciais da cadeia são correlacionadas e algum tempo é necessário para que a cadeia 'esqueça' os valores iniciais e atinja a distribuição equilíbrio, da qual as amostras serão usadas para inferência. A velocidade que se dá a convergência depende da complexidade do modelo, da correlação a *posteriori* dos parâmetros e da autocorrelação entre amostras sucessivas (CARDOSO, 2008).

Outro aspecto importante na inferência a *posteriori* é a estimação da variância de Monte Carlo para os estimadores, que indica quanto que os

valores obtidos para essas estatísticas podem variar em função da dimensão finita da cadeia. Assim, a acurácia das estimativas pode ser dada em função do erro de Monte Carlo (SORENSEN et al., 1994), que é estimado mediante o cálculo da variância das amostras retiradas para cada componente de variância e dividindo-se essa variância pelo número de amostras. Dessa forma, a raiz quadrada desse valor refere-se a aproximação do desvio-padrão do erro associado com o tamanho da cadeia de Gibbs.

Distribuição a *priori*

A interpretação para probabilidades se refere às probabilidades subjetivas, nas quais é levado em consideração o 'grau de confiança' de um determinado indivíduo. As técnicas Bayesianas utilizam probabilidades subjetivas medindo graus de confiança dos valores dos parâmetros desconhecidos (θ). Essas probabilidades subjetivas são usadas para definir o que é chamado de distribuição a *priori* para o parâmetro (θ). Então, ao trabalhar com métodos Bayesianos, tem-se o parâmetro desconhecido como uma variável aleatória, cuja distribuição a *priori* (antes de se ter os valores observados da amostra) é conhecida. Essa distribuição a *priori* resume o grau de confiança subjetivo sobre o valor desconhecido do parâmetro (SORENSEN; GIANOLA, 2002).

Crítica comum feita a aproximação Bayesiana, é que a escolha da distribuição a *priori* é relativamente subjetiva. Este questionamento está relacionado com o fato que, em alguns casos, a distribuição a *posteriori* é muito sensível à escolha da distribuição a *priori*. Neste caso, dois pesquisadores usando o mesmo conjunto de dados podem chegar a diferentes conclusões se usarem diferentes distribuições a *priori*.

Quando o conhecimento prévio sobre o fenômeno estudado é vago, ou praticamente inexistente, a densidade a *priori* é denominada não informativa, plana, ou *flat*, possuindo distribuição uniforme (Figura 2), representada como:

$f(\theta) \propto \text{constante}$

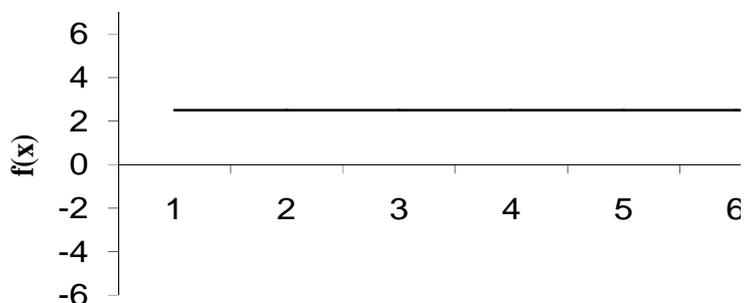


Figura 2 – Representação gráfica de distribuição com *priori* não informativa

Embora controversas, a maioria dos argumentos da inferência frequêntista, podem ser derivados pelo teorema de Bayes, usando uma *priori* não informativa (GIANOLA; FERNANDO, 1986).

Por outro lado, quando se possui conhecimento prévio amplo ou preciso sobre o fenômeno estudado, a densidade a *priori* passa a ser denominada de informativa, aguda, ou *sharp*. Neste caso, a maioria da densidade desta função, está a vizinhança de sua moda (Figura 3).

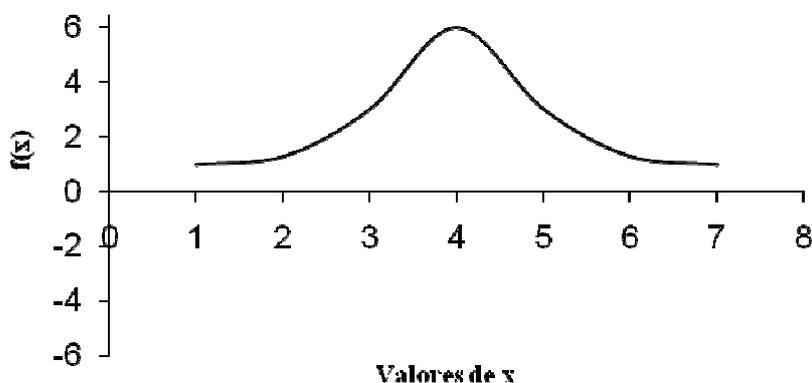


Figura 3 – Representação gráfica de distribuição com *priori* informativa

Wang et al. (1993), avaliaram o efeito de quatro níveis de informação *a priori* em dados sob seleção e ilustraram como informações externas podem contribuir com as informações geradas pelos dados. Os autores observaram que as estimativas foram mais precisas quando maiores níveis de informação *a priori* foram utilizados, demonstrando que a análise Bayesiana pode ser empregada para incorporar informações externas em uma situação prática de melhoramento animal.

Nessa mesma linha de raciocínio, Carneiro Júnior et al. (2005), verificaram que a inclusão de informação *a priori* proporciona melhores resultados para estimação dos componentes de variância, sobretudo em populações pequenas. Concluíram também que à medida que o tamanho da amostra aumenta a importância da informação *a priori* decresce e com isso a predição dos valores genéticos é menos influenciada pela informação *a priori*. Porém, a estimação dos componentes de variância, de modo que a escolha dos melhores indivíduos é mais prejudicada em populações com pequeno número de animais, mesmo quando maiores níveis de informação são disponíveis.

Considerações Finais

Os procedimentos aplicados para características nas quais a variação contínua é assumida não são adequados para aquelas com variação não contínua, com isso o desenvolvimento de metodologias específicas para análises de dados categóricos veio, ao encontro das necessidades dos pesquisadores, pois auxilia o estudo de características de dados categóricos, viabilizando sua utilização na avaliação genética do rebanho.

Por fim, pretende-se deixar claro que o propósito principal deste trabalho constitui em oferecer linhas de pensamento que, provavelmente, funcionem como um recurso adicional no processo de análise para características com distribuição não-normal.

Referencias Bibliográficas

- BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v. 79, p. 2023-2046, 2001.
- BROOKS, S. P.; ROBERTS, G. O. Convergence assessment techniques for Markov chain Monte Carlo. **Statistics and Computing**, v. 8, p.319-335, 1998.
- CARDOSO, F. F. **Manual de utilização do programa INTERGEN**: versão 1.0 em estudos de genética quantitativa animal. Bagé: Embrapa Pecuária Sul, 2008.
- CARNEIRO JÚNIOR, J. M.; ASSIS, G. M. L.; EUCLYDES, R. F. et al. Influência da informação a *priori* na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v. 34, p. 1905-1913, 2005.
- COWLES, M. K.; CARLIN, B. P.. Markov chain Monte Carlo convergence diagnostics: a comparative review. **Journal of the American Statistical Association**, v. 91, p. 883-904, 1996.
- FALCONER, D. S. The inheritance of liability to certain diseases estimated from the incidence from relatives. **Annals Human Genetics**, v. 29, p. 51-76, 1965.
- FARIA, C. U.; MAGNABOSCO, C. U.; LÔBO, R. B. et al. Inferência bayesiana e sua aplicação na avaliação genética de bovinos da raça Nelore: revisão bibliográfica. **Ciência Animal Brasileira**, v. 8, p. 75-86, 2007.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.721-741, 1984.
- GEYER, C. J. Practical Markov chain Monte Carlo. **Statistical Science**, v. 7, p. 473-511, 1992.
- GIANOLA, D. Genetic evaluation of animals for traits with categorical responses. **Journal Animal Science**, v. 51, p. 1272-1276, 1980.
- GIANOLA, D. Theory and analysis of threshold characters. **Journal Animal Science**. v. 54, p. 1079-1081, 1982.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal Animal Science**, v.63, p.217-244, 1986.
- GIANOLA, D.; FOULLEY, J. L. Sire evaluation for ordered categorical data with a threshold model. **Genetics Selection Evolution**, v. 15, p. 201-224, 1983.
- HASTING, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97-109, 1970.
- HOESCHELE, I.; GIANOLA, D. Bayesian versus maximum quaslikelihood methods for sire evaluation with categorical data. **Journal of Dairy Science**, v.72, p.1569-1577, 1988.
- JENSEN, J. Sire evaluation for type traits with linear and non-linear procedures. **Livestock Production Science**, v. 15, p. 165-171, 1986.
- JOHNSON, V. E. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. **Journal of the American Statistical Association**, v. 91, p. 154-166, 1996.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. 2 ed. Wallingford: CABI Publishing, 2005.
- NOGUEIRA, D. A.; SÁFADI, T.; BEARZOTI, T.; et al. Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e Agrotecnologia**, Edição Especial, p.1614-1624, 2003.

OLESEN, I.; PEREZ-ENCISO, M.; GIANOLA, G. et al. A comparison of normal and nonnormal mixed models for number of lambs Born in Norwegian sheep. **Journal Animal Science**, v. 72, p. 1166-1173, 1994.

SORENSEN, D. A., WANG, C. S., JENSEN, J. et al. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetics Selection Evolution**, v. 26, 333-360, 1994.

SORENSEN, D. A.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**: statistics for biology and health, New York: Springer-Verlag, 2002.

SOUSA, W. H.; PEREIRA, C. S.; BERGMANN, J. A. G. et al. Estimativas de componentes de variância e de parâmetros genéticos para características de reprodução por intermédio de modelos lineares e de limiar. **Revista Brasileira de Zootecnia**. v. 29, p. 2237-2247, 2000.

SOUZA, W. H. **Aplicação de modelos lineares e não-lineares em características de reprodução, sobrevivência e crescimento em ovinos da raça Santa Inês**. 1997. 131 f. Tese (Doutorado) - Universidade Federal de Minas Gerais, Belo Horizonte, 1997.

THOMPSON, R. Sire evaluation. **Biometrics**, v. 35, p. 339-353, 1979.

TURNER, H. N.; YOUNG, S. S. Y. Genetic improvement of reproduction rate in sheep. *Animal Breeding Abstr.*, v. 37, p. 545-563, 1969.

VAN TASSEL, C. P., CASELLA, G., POLLAK, E. J. et al. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v. 78, p. 678-692, 1995.

WANG, C. S.; RUTLEDGE, J. J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. **Genetics Selection Evolution**, v. 25, p. 41-62, 1993.

WRIGHT, S. An analysis of variability in number of digits in an inbred strain of guinea pigs. **Genetics**, v. 19, p. 506-536, 1934.